5/

ENTERED

PCT09

RAW SEQUENCE LISTING DATE: 03/19/2003
PATENT APPLICATION: **US/09/601,278** TIME: 07:16:55

Input Set : **A:\p-1476-US.ST25.txt**
Output Set: **N:\CRF4\03192003\I601278.raw**

```
      3 <110> APPLICANT: Yissum Research and Development of the Hebrew University of
Jerusalem.
      4       Linial, Michal
      5       Linial, Nathan
      6       Tishbi, Naftali
      7       Yona, Golan
      9 <120> TITLE OF INVENTION: An automatic method of classifying molecules
     11 <130> FILE REFERENCE: P-1476-US
     13 <140> CURRENT APPLICATION NUMBER: 09/601,278
     14 <141> CURRENT FILING DATE: 1999-01-29
     16 <150> PRIOR APPLICATION NUMBER: 60/072,977
     17 <151> PRIOR FILING DATE: 1999-01-29
     19 <160> NUMBER OF SEQ ID NOS: 4
     21 <170> SOFTWARE: PatentIn version 3.1
     23 <210> SEQ ID NO: 1
     24 <211> LENGTH: 122
     25 <212> TYPE: PRT
     26 <213> ORGANISM: Saccharomyces cerevisiae
     28 <400> SEQUENCE: 1
     30 Lys Asp Lys Ile Val Leu Asp Val Gly Cys Gly Thr Gly Ile Leu Ser
     31 1               5                  10                  15
     34 Met Phe Ala Ala Lys His Gly Ala Lys His Val Ile Gly Val Asp Met
     35             20                  25                  30
     38 Ser Ser Ile Ile Glu Met Ala Lys Glu Leu Val Glu Leu Asn Gly Phe
     39         35                  40                  45
     42 Ser Asp Lys Ile Thr Leu Leu Arg Gly Lys Leu Glu Asp Val His Leu
     43     50                  55                  60
     46 Pro Phe Pro Lys Val Asp Ile Ile Ile Ser Glu Trp Met Gly Tyr Phe
     47 65                  70                  75                  80
     50 Leu Leu Tyr Glu Ser Met Met Asp Thr Val Leu Tyr Ala Arg Asp His
     51                 85                  90                  95
     54 Tyr Leu Val Glu Gly Gly Leu Ile Phe Pro Asp Lys Cys Ser Ile His
     55             100                 105                 110
     58 Leu Ala Gly Leu Glu Asp Ser Gln Tyr Lys
     59         115                 120
     62 <210> SEQ ID NO: 2
     63 <211> LENGTH: 123
     64 <212> TYPE: PRT
     65 <213> ORGANISM: haemophilus sp
     67 <400> SEQUENCE: 2
     69 Lys Gly Lys Lys Leu Leu Asp Leu Gly Cys Gly Thr Gly Gly His Leu
     70 1               5                  10                  15
     73 Gln Leu Tyr Leu Glu Arg Gly Ala Ala Lys Val Ile Gly Thr Asp Leu
     74             20                  25                  30
```

RAW SEQUENCE LISTING DATE: 03/19/2003
PATENT APPLICATION: US/09/601,278 TIME: 07:16:55

Input Set : A:\p-1476-US.ST25.txt
Output Set: N:\CRF4\03192003\I601278.raw

```
77 Ser Glu Lys Met Leu Glu Gln Ala Glu Lys Asp Leu Gln Lys Cys Gly
78         35                  40                  45
81 Gln Phe Ser Gly Arg Phe Ser Leu Tyr His Leu Pro Ile Glu Lys Leu
82     50                  55                  60
85 Ala Glu Leu Pro Glu Ser His Phe Asp Val Ile Thr Ser Ser Phe Ala
87 65                  70                  75                  80
90 Phe His Tyr Ile Glu Asn Phe Pro Thr Leu Leu Ser Thr Ile His Asp
91                 85                  90                  95
94 Lys Leu Ser Ser Asn Gly Thr Leu Ile Phe Ser Gln Glu His Pro Ile
95             100                 105                 110
98 Thr Thr Cys His Lys Glu Gly Glu Arg Trp Glu
99         115                 120
102 <210> SEQ ID NO: 3
103 <211> LENGTH: 210
104 <212> TYPE: PRT
105 <213> ORGANISM: Gluconacetobacter xylinus
107 <400> SEQUENCE: 3
109 Pro Leu Pro Asp Asn Val Asp Asp Trp Pro Thr Val Asp Ile Phe Ile
110 1               5                  10                  15
113 Pro Thr Tyr Asp Glu Gln Leu Ser Ile Val Arg Leu Thr Val Leu Gly
114             20                  25                  30
117 Ala Leu Gly Ile Asp Trp Pro Pro Asp Lys Val Asn Val Tyr Ile Leu
118         35                  40                  45
121 Asp Asp Gly Val Arg Pro Glu Phe Glu Gln Phe Ala Lys Asp Cys Gly
122     50                  55                  60
125 Ala Leu Tyr Ile Gly Arg Val Asp Ser Ser His Ala Lys Ala Gly Asn
126 65                  70                  75                  80
129 Leu Asn His Ala Ile Lys Arg Thr Ser Gly Asp Tyr Ile Leu Ile Leu
130                 85                  90                  95
133 Asp Gly Asp His Ile Pro Thr Arg Ala Phe Leu Gln Ile Ala Met Gly
134             100                 105                 110
137 Trp Asn Val Ala Asp Arg Lys Ile Ala Leu Met Gln Thr Pro His His
138         115                 120                 125
141 Phe Tyr Ser Pro Asp Pro Phe Gln Arg Asn Leu Ala Val Gly Tyr Arg
142     130                 135                 140
145 Thr Pro Pro Glu Phe Asn Leu Phe Tyr Gly Val Ile Gln Asp Gly Asn
146 145                 150                 155                 160
149 Asp Phe Trp Asp Ala Thr Phe Phe Cys Gly Ser Cys Ala Ile Leu Arg
150                 165                 170                 175
153 Arg Glu Ala Ile Glu Ser Ile Gly Gly Phe Ala Val Glu Thr Val Thr
154             180                 185                 190
157 Glu Asp Ala His Thr Ala Leu Arg Met Gln Arg Arg Gly Trp Ser Thr
158         195                 200                 205
161 Ala Tyr
162     210
165 <210> SEQ ID NO: 4
166 <211> LENGTH: 201
167 <212> TYPE: PRT
168 <213> ORGANISM: Rhodococcus sp.
```

RAW SEQUENCE LISTING DATE: 03/19/2003
PATENT APPLICATION: **US/09/601,278** TIME: 07:16:55

Input Set : **A:\p-1476-US.ST25.txt**
Output Set: **N:\CRF4\03192003\I601278.raw**

170 <400> SEQUENCE: 4
172 Pro Arg Pro Met Ser Thr Pro Ser Ala Ala Asp Val Thr Val Val Ile
173 1 5 10 15
176 Pro Val Lys Asp Asn Gln Ala Gly Val Glu Arg Leu Leu Pro Val Leu
177 20 25 30
180 Asp Lys Leu Thr Val Ile Val Val Asp Asp Gly Ser Glu Val Pro Val
181 35 40 45
184 Glu Pro Arg Arg Ala Cys Pro Gly Thr Gly Thr Ile Thr Val Val Arg
185 50 55 60
188 His Glu Ser Ala Arg Gly Pro Ser Ala Ala Arg Asn Ser Gly Leu Arg
189 65 70 75 80
192 Ser Ala Gln Thr Arg Phe Val Ala Phe Leu Asp Ser Asp Val Ile Pro
193 85 90 95
196 Arg Ala Gly Trp Leu Glu Leu Met Leu Gly His Phe Ser Asp Pro Gly
197 100 105 110
200 Val Ala Leu Val Ala Pro Arg Ile Val Ala Leu Asp Pro Tyr Gly Thr
201 115 120 125
204 Ala Leu Ala Arg Tyr Glu Asn Met Arg Ser Ser Leu Asp Leu Gly Arg
205 130 135 140
208 Lys Glu Ala Ala Val Lys Ser Gly Ser Pro Val Ala Tyr Val Pro Ser
209 145 150 155 160
212 Ala Ala Val Ile Val Arg Arg Asp Val Ala Leu Glu Cys Asn Gly Phe
213 165 170 175
216 Asp Glu Ser Leu Glu Val Ala Glu Asp Val Asp Phe Cys Trp Arg Leu
217 180 185 190
220 Gln Ala Ala Gly Trp Arg Leu Arg Tyr
221 195 200

VERIFICATION SUMMARY
PATENT APPLICATION: **US/09/601,278**

DATE: 03/19/2003
TIME: 07:16:56

Input Set : **A:\p-1476-US.ST25.txt**
Output Set: **N:\CRF4\03192003\I601278.raw**